应用预测最小二乘准则的 AR 模型判阶方法 [1]

石 星　　戴庆芬*　　李乐民

(电子科技大学信息研究所 成都 610054)
*(电子部十所 成都 610039)

摘 要 本文通过多种 AR 模型的判阶准则的比较提出了应用最小二乘预测误差的滑窗预测最小二乘 (Sliding Window Predictive Least Squares ,SWPLS) 判阶准则. 采用这种准则的主要优点除了准确的判阶性能外, 对于时变 AR 模型具有良好的跟踪特性, 同时算法容易实现在线实时处理. 文中主要对时变模型参数和时变模型阶数的多种情况进行了判阶模拟, 验证了文中提出的滑窗最小二乘预测判阶准则的有效性.

关键词 模型判阶, 滑窗数据, 格形结构

中图号 TN911.7

1 前 言

采用 AR 模型进行实际观测序列的拟合是信号处理的多种应用中经常采用的方法. 对于有限长度的序列这种方法具有高分辨的谱特性. 但实施 AR 模型拟合的一个关键问题是模型阶数的有效判决. 阶数过大将引起谱线分裂, 即出现伪峰, 阶数过小将平滑谱特性. 关于有效的模型阶数判决仍是 AR 模型研究中有待解决的问题 [1] . 针对 AR 模型的阶数判决已有多种准则, 包括 AIC(Akaike Informatoin Criterion) 准则和 MDL(Minimum Description Length) 准则等. 在这些准则中大多需要满足数据为高斯分布或渐近高斯分布的条件且需经过二次过程: 首先估计最佳参数, 然后进行预测误差计算. 显然这不适合于在线实现. 在下节中将介绍的准则如 SVD(Singular Value Decomposition) 准则, PDC (Predictive Density Criterion) 准则和 FSD(Fast Subspace Decomposition) 准则计算量都很大, 实时处理难度较大.

Mati Wax 在文献 [2] 中提出了预测最小二乘 (Predictive Least Square, PLS) 的 AR 模型判阶方法. 这种方法的实现无需数据的高斯分布假设限制且仅需一次过程, 在阶数判决的同时又可得到模型参数的估计. PLS 方法被证明具有阶数选择的一致性 [2] . 实现 PLS 的两种主要方法是前加窗最小二乘预测 (PWPLS) 方法和记忆增长协方差最小二乘预测 (GMCPLS) 方法. 对于高阶的模型阶数判决后者比前者更加有效 [3] . 然而实际应用中经常遇到短数据和时变模型的情况, 有时以满足局部平稳的条件出现. 针对这种参数时变模型的判阶算法目前的研究仍然很少. 在模型阶数和模型参数都具时变性的情况下, 有效的实时判阶更加困难. 本文在对目前已有的模型阶数判决方法进行全面比较分析的基础上提出了滑窗预测最小二乘 (SWPLS) 的 AR 模型判阶方法. 这种判阶方法不但有广泛的适应性且计算量居于 PWPLS 和 GMCPLS 之间, 可针对时变的程度调节数据窗长度以对时变模型阶数和参数进行有效跟踪. 计算机模拟的结果验证了算法对时变模型的良好判阶特性. 全文将首先介绍 AR 模型判决的几种典型算法. 然后重点研究预测最小二乘准则下的格形实现判阶方法, 提出有效的 SWPLS 方法, 并结合实验结果进行性能分析, 证明算法的有效性.

[1] 1996-03-20 收到, 1997-01-06 定稿

2 AR 模型的判阶准则 [4]

采用 AR 模型对时间序列 $\{x(n), 1 \le n \le N\}$ 进行拟合的模型结构为

$$x(n) = \sum_{i=1}^{p} a(i)x(n-i) + n(i), \tag{1}$$

式中 p 为模型阶数, $n(i)$ 为零均值方差为 σ_n^2 的高斯白噪声序列. 建立模型的关键是通过各种谱估计算法进行各阶参数的估计, 另一主要问题就是对阶数 p 的判决. 下面先介绍几种典型的和新近出现的判阶准则.

2.1 AIC 准则

AIC 准则的阶数估计算法可描述为

$$\hat{p} = \arg\left\{\min_{k\in\{0,1,\cdots,p_{\max}\}} \left(\ln \hat{\sigma}_k^2 + 2k/N\right)\right\}, \tag{2}$$

式中 $\arg\{\cdot\}$ 表示求取括号内条件下的最佳 k 值, $\hat{\sigma}_k^2$ 为激励白噪声方差的最大似然估计, $p_{\max}$ 为最大可选阶数.

2.2 MDL 准则

MDL 准则可表示为

$$\hat{p} = \arg\left\{\min_{k\in\{0,1,\cdots,p_{\max}\}} \left(\ln \hat{\sigma}_k^2 + (k/N)\ln N\right)\right\}. \tag{3}$$

AIC 准则不具备阶数估计的一致性, MDL 对其不足进行了改进. 上述两种准则为 AR模型判阶中较典型的准则. 除判阶的非一致性外, 它们的缺点还包括应用于非高斯数据仍有问题, 同时判阶中的二次过程使它们仅适宜于脱线计算的场合.

2.3 SVD 的判阶准则

SVD 的判阶准则利用观察过程的数据矩阵, 通过确定矩阵的有效秩来决定 AR 模型的阶数. 数据矩阵 $X \in R^{l\times m}$,

$$\begin{aligned}
&X = [\boldsymbol{x}(N-1), \boldsymbol{x}(N-2), \cdots, \boldsymbol{x}(N-m)],\\
&\boldsymbol{x}(N-k) \in R^l, \quad k = 1, 2, \cdots, m,\\
&\boldsymbol{x}(N-k) = [x(N-k), x(N-k-1), \cdots, x(N-k-l+1)]^T,
\end{aligned} \tag{4}$$

T 表示转置. 设 X 的奇异值为 $\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_m$, 模型的阶数由关系: $\alpha_p > \delta \ge \alpha_{p+1}$ 决定, 门限值由下述关系决定: $\delta = \sqrt{c(l)}\hat{\sigma}_u$, 其中 $c(l)$ 为自由度为 l 的 χ^2 分布的百分位值, $\hat{\sigma}_u$ 为激励噪声的均方差估计值.

2.4 FSD 的判阶方法

FSD 方法用于数据矩阵或数据协方差矩阵的信号子空间维数判决. 由于原理相似可直接用于 AR 模型的阶数判决. 当数据矩阵 $X \in R^{N\times M}$, $\theta_k^{(m)}$ 为矩阵 $1/N(X^TX)$ 经 m 步 Lanczos 快速特征值分解算法的第 k 个 Rayleigh-Ritz(RR) 值. FSD 的判决式为

$$\varphi_{\hat{d}} = \sqrt{N}[\log(\hat{\sigma}_{\hat{d}}/\hat{\sigma}_{\hat{d}+1})], \tag{5}$$

式中 $\hat{\sigma}_j=\frac{1}{M-j}\left(\left\|\frac{1}{\sqrt{N}}X^TX\right\|_F^2-\sum_{k=1}^{j}(\theta_k^{(m)})^2\right)$. 选择门限 $r_d c(N)$ 为 $O(\sqrt{N})$ 与 $O(\sqrt{\log\log N})$ 之间的值, 则当 $\varphi_{\hat{d}}\le r_d c(N)$ 时, $\hat{d}$ 确定为正确的阶数估计, 否则 $\hat{d}$ 的估计不成立. 令 $\hat{d}=\hat{d}+1$, 重新判别, 当 $\hat{d}<m-2$, 重复上述过程, 直到正确的阶数估计.

2.5 PDC 判决方法

PDC 准则是以可选阶数的整体误差概率来决定最佳的阶数选取. $x(n)$ 的预测概率密度为 $f(x(n)/\boldsymbol{x}(1,n-1),k)$, $\boldsymbol{x}(1,n-1)=[x(1),\cdots,x(n-1)]$, k 为可选的阶数, 则 PDC 准则表示为

$$\hat{p}=\arg\left\{\min_{k\in\{0,1,\cdots,p_{\max}\}}\left[\sum_{n-1}^{N}-\ln(f(x(n)/\boldsymbol{x}(1,n-1),k))\right]\right\} \tag{6}$$

文献 [4] 中详细说明了 $f(x(n)/\boldsymbol{x}(1,n-1),k)$ 的计算方法.

上述三种判阶准则都仅需一次数据过程. SVD 对于模型阶数较小时, 判阶特性较好. PDC 对于短数据的情况有好的判阶特性. FSD 更适合于宽带数据加噪声的情况, 且判阶特性优于 AIC 和 MDL 方法. 这三种方法共同存在的一个不足是运算量还显得太大. 尤其是 PDC 方法, 其计算量是下节将介绍的 GMCPLS 方法的十倍, 而判阶性能上对高的阶数后者却优于前者.

3 滑窗最小二乘预测 (SWPLS) 的判阶方法

在时变模型的情况下, 要求对模型阶数和模型参数的变化进行跟踪, 这需要丢弃旧数据换以新数据. 一种方法是采用加权方法使新数据有更大的权值. 但加权将使数据协方差矩阵发生变化, 同时旧数据的影响不可能完全消失. 这样得到的预测误差不可能是新数据的精确预测误差从而影响判阶特性. 采用滑窗数据方式可以完全避免旧数据的影响, 同时根据数据时变的程度可调节窗长达到最佳的跟踪性能.

对于阶数为 p 的模型, 真实预测误差表示为

$$\hat{e}(p,n)=x(n)-\boldsymbol{a}^T(p,n-1)\boldsymbol{x}(p,n-1), \tag{7}$$

$\boldsymbol{a}(p,n-1)=[a(1,n-1),\cdots,a(p,n-1)]^T$ 为 p 阶预测滤波的参数矢量. SWPLS 的判阶准则可表示为

$$\hat{p}=\arg\left\{\min_{k\in\{0,1,\cdots,p_{\max}\}}\left[\frac{1}{W}\sum_{n=N-W+1}^{N}\hat{e}^2(k,n)\right]\right\}, \tag{8}$$

式中 W 为滑窗长度. 当 $W=N$ 时, 上述准则为前加窗的 PWPLS 准则. 当窗长 $W=N-2p_{\max}$ 时则成为 GMCPLS 准则. 这说明 SWPLS 有很大的灵活性, 可方便地变形为 PWPLS 或 GMCPLS 准则. 但在一般情况下 W 应根据模型的时变程度进行调节.

在局部平稳区域相同于文献 [2] 中的三个假设仍然有效. 这三个假设是:

(1) 样本数据由 p 阶非变化 AR 过程产生:

$$x(n)=\sum_{i=1}^{p}a(i)x(n-i)+u(i). \tag{9}$$

(2) 特征方程

$$z^p-a(1)z^{p-1}-,\cdots,-a(p)=0 \tag{10}$$

的根的绝对值小于 1．

(3) 对于足够大的 n 满足：

$$\|C_{p,n}C_p^{-1} - I_p\|_2 < 1, \tag{11}$$

其中 $C_{p,n} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}(p,n-i)\boldsymbol{x}^T(p,n-i)$, $C_p = E[\boldsymbol{x}(p,n)x^T(p,n)]$, $\boldsymbol{x}(p,n) = [x(n-1),\cdots,x(n-p)]^T$．在满足上述三个假设条件下，应用 SWPLS 的判阶准则，则阶数估计满足一致性条件，即当样本数 n 足够大时，

$$P_r[\hat{p} = p] = 1. \tag{12}$$

然而应用滑窗由于滤波失调增加将使滤波误差有一定程度增大．误差的渐近值为

$$\eta = e_{\min}^2(1 + p/W) = [D(x) - \boldsymbol{a}^T C_p \boldsymbol{a}](1 + p/W), \tag{13}$$

式中 $D(x)$ 为数据样本方差， $\boldsymbol{a} = [1(1),\cdots,a(p)]^T$ 为模型参数矢量．由于是 SWPLS 准则内误差的一致比较 (相对预测误差的最小渐近值)，因而失调增加对判阶影响不会很大，在跟踪性能得以满足的条件下，尽可能加大窗长，将有助于减小失调．同时考虑模型阶数的时变性，阶数的判决应根据不同样本区间滤波误差的相对最小而加以判决． SWPLS 具有与 PWPLS 和 GMCPLS 相似的阶数递推格形实现方式，在这一过程中前后预测误差同时最小化，后向预测误差由于 Schmit 正交化过程而相互正交．

4　模拟结果和分析

这里通过与 PWPLS 和 GMCPLS 的比较来说明 SWPLS 准则的性能．关于 PWPLS 和 GMCPLS 与其它判阶方法的比较，文献 [3] 和文献 [4] 中有详细的结果．模拟实验在 Pentium586 微机上用 Borland C++3.0 语言完成．模拟选择了三个比较典型的 2 阶、 4 阶和 6 阶 AR 模型 [1,4]，如下所示：

$$\begin{aligned}
&\mathrm{AR}(2) \quad x(n) = 1.8x(n-1) - 0.95x(n-2) + u(n),\\
&\mathrm{AR}(4) \quad x(n) = 2.76x(n-1) - 3.81x(n-2) + 2.65x(n-3) - 0.92x(n-4) + u(n),\\
&\mathrm{AR}(6) \quad x(n) = 4.23x(n-5) - 8.19x(n-2) + 9.34x(n-3) - 6.64x(n-4)\\
&\qquad\qquad + 2.78x(n-5) - 0.53x(n-6) + u(n).
\end{aligned} \tag{14}$$

性能的比较按命中率来评价．这里命中率是指在一批独立实验中按准则估计的模型阶数等于真实模型阶数的实验次数．总的实验次数为本 100，取的样本点数为本 120，最大的可选阶数 $p_{\max}$ 取为 8．为验证 SWPLS 对时变模型的跟踪性能，特别进行了从不同模型间变化的判阶实验，变化前和变化后的模型分别取相等的样点，最后评判对变化后模型的命中率． SWPLS 的递推格形关系式列于附录中．模拟比较结果如表 1 和表 2 所示．

表 1 平稳过程判阶命中率比较

过程	准则			
	GMCPLS	PWPLS	SWPLS W=40	SWPLS W=50
AR(2)	88	87	80	83
AR(4)	89	80	77	81
AR(6)	91	79	76	84

表 2 时变过程判阶命中率比较

变化过程变化过程	准则			
	GMCPLS	PWPLS	SWPLS W=40	SWPLS W=50
AR(2)→AR(4)	65	58	75	80
AR(4)→AR(6)	63	53	76	79
AR(4)→AR(2)	68	65	77	82

从表 1 看出对于时不变的 AR 模型相比之下 GMCPLS 准则效果最好，PWPLS 和 SWPLS 相比性能略差. SWPLS 的窗长增加将使判阶性能提高，特别对高阶模型更加明显. 表 2 是时变模型的模拟结果. 显然 SWPLS 准则的优点在这时得到了体现，即使与 GMCPLS 比较命中率平均高出十多个百分点. 调节 W 同样使判阶性能改变. 一般情况下适当加长 W 效果会显得更好些. 但 W 的上限一般确定为不同平稳区间的平均值. 实用中 W 的调节应取得跟踪特性和判阶性能上的平衡.

5 结 论

全文在综合比较分析典型和新近出现的 AR 模型判阶准则的基础上，提出了应用滑窗格形递推自适应滤波误差在时变模型参数和时变模型阶数的情况下，进行 AR 模型判阶的滑窗预测最小二乘 (SWPLS) 准则. 说明了阶数判决在局部平稳区的一致性，模拟实验的结果证明了这种判决方法的有效性. 由于可根据时变模型的变化情况进行窗长调节，因而判阶方法具有很好的适应性. 下一步准备采用具有更好数值特性的 QR 分解滑窗递推 LS 方法进行滤波误差统计，估计会有更好的效果.

附 录

这里参照文献 [2,3,6] 列出 SWPLS 算法步骤.

初始化:

$$x(n) = 0, \quad n < 0;$$

$$e_0^f(n) = e_0^b(n) = x(n);$$

$$\varepsilon_0^f(n) = \varepsilon_0^b(n) = \varepsilon_0^f(n-1) + x^2(n); \quad \varepsilon_0^f(0) = x^2(0);$$

$$\Delta_p(n) = 0, \quad n < p; \quad \Delta_{p,w-1}(n) = 0, \quad n < p;$$

$$r_0(n) = \xi_0(n) = 1;$$

$$d_0(0) = x(0); \quad d_p(0) = 0, \quad p > 1;$$

$$g_0(n) = x(x(n - W + 1);$$

对 $p = 0 \rightarrow \min(k, p_{\max})$, $n = 1, \cdots$;

$$\Delta_{p+1}(n) = \Delta_{p+1,W-1}(n-1) + [e_p^b(n-1)e_p^f(n)]/[r_p(n-1)],$$

$$\Delta_{p+1,W-1}(n) = \Delta_{p+1}(n) - [g_p(n)d_p(n-1)]/[\xi_p(n-1)],$$

$$e_{p+1}^f(n) = e_p^f(n) - K_{p+1}^b(n)e_p^b(n-1); \quad K_{p+1}^f(n) = \Delta_{p+1}(n)/(\varepsilon_p^f(n)),$$

$$e_{p+1}^b(n) = e_p^b(n) - K_{p+1}^f(n)e_p^f(n); \quad K_{p+1}^b(n) = \Delta_{p+1}(n)/(\varepsilon_p^b(n-1)),$$

$$\varepsilon_{p+1}^f(n) = \varepsilon_p^f(n) - K_{p+1}^b(n)\Delta_{p+1}(n-1); \quad \varepsilon_{p+1}^b(n) = \varepsilon_p^b(n-1) - K_{p+1}^f(n)\Delta_{p+1}(n);$$

$$r_{p+1}(n) = r_p(n) - e_p^b(n)d_p(n)/(\varepsilon_p^b(n)); \quad \xi_{p+1}(n) = \xi_p(n) - d_p^2(n)/(\varepsilon_p^b(n));$$

$$d_{p+1}(n) = d_p(n-1) - \Delta_{p+1}(n)g_p(n)/(\varepsilon_p^f(n));$$

$$g_{p+1}(n) = g_p(n) - \Delta_{p+1}(n)d_p(n-1)/(\varepsilon_p^b(n-1)).$$

准则需要的真实误差为 $e(k,n) = e_k^f(n)/r_k(n-1)$. PWPLS 和 GMCPLS 的算法步骤可参考文献 [2,3] .

参 考 文 献

[1] Kay S M. Modern Spectral Estimation: Theory and Application. Englewood Cliffs: Prentice Hall, 1988 Chapter 6.

[2] Wax M. Order selection for AR models by predictive least squares. IEEE Trans. on ASSP, 1988, ASSP-36(4): 581–588.

[3] Nandi A K, Chambers J A. New lattice realization of the predictive least squares order selection criterion. IEE Proc.-F , 1991, 138(6): 545–550.

[4] Dickie T R, Nandi A K. A comparative study of AR order selection methods. Signal Processing, 1994, 40: 239–255.

[5] Guanghan Xu, Kailath T. Fast subspace decomposition. IEEE Trans. on SP, 1994, SP-42(3): 539–551.

[6] Simon Hakay. Adaptive Filtering Theory. Englewood Cliffs: Pretice Hall, 1991, Chapter7.

STUDY OF AR MODEL ORDER SELECTION APPLYING CRITERION OF PREDICTIVE LEAST SQUARES

Shi Xing　Dai Qingfen*　Li Lemin

(*University of Electronic Science and Technology of China, Chengdu* 610054)

*(*10th Institute of Electronic Industry Ministry, Chengdu* 610039)

Abstract A criterion called Sliding Window Predictive Least-Squares (SWPLS) order selection criterion is presented in this paper by comparison of several typical criteria of AR model order selection. In addition to property of fine order selection the criterion has advantage of good tracking property and can be easily implemented on-line in real time. The effectiveness of SWPLS criterion is verified by simulation experiment of order selection for models of time-varying parameter and time-varying order.

Key words Order selection of model, Sliding window data, Lattice structure

石　星：　男，1961 年生，博士生，主要从事信号和信息处理及其应用的研究.

戴庆芬：　女，1937 年生，研究员，主要从事通信及雷达的信号和信息处理研究.

李乐民：　男，1932 年生，教授，博士生导师，主要从事数字信号传输和处理、光纤通信和通信网等方面的研究.